



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### Identification of a Conserved RNA Replication Element (cre) within the 3Dpol-Coding Sequence of Hepatoviruses

**Citation for published version:**

Yang, Y, Yi, M, Evans, DJ, Simmonds, P & Lemon, SM 2008, 'Identification of a Conserved RNA Replication Element (cre) within the 3Dpol-Coding Sequence of Hepatoviruses', *Journal of Virology*, vol. 82, no. 20, pp. 10118-10128. <https://doi.org/10.1128/JVI.00787-08>

**Digital Object Identifier (DOI):**

[10.1128/JVI.00787-08](https://doi.org/10.1128/JVI.00787-08)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Peer reviewed version

**Published In:**

Journal of Virology

**Publisher Rights Statement:**

Copyright © 2008, American Society for Microbiology. All Rights Reserved.

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



## Identification of a Conserved RNA Replication Element (*cre*) within the 3D<sup>pol</sup>-Coding Sequence of Hepatoviruses<sup>∇</sup>

Yan Yang,<sup>1†‡</sup> MinKyung Yi,<sup>1‡</sup> David J. Evans,<sup>2</sup> Peter Simmonds,<sup>3</sup> and Stanley M. Lemon<sup>1\*</sup>

Center for Hepatitis Research, Institute for Human Infections and Immunity, and the Department of Microbiology and Immunology, University of Texas Medical Branch, Galveston, Texas 77555-1073<sup>1</sup>; Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, United Kingdom<sup>2</sup>; and Virus Evolution Group, Centre for Infectious Diseases, University of Edinburgh, Summerhall, Edinburgh EH9 1QH, United Kingdom<sup>3</sup>

Received 12 April 2008/Accepted 29 July 2008

**Internally located, *cis*-acting RNA replication elements (*cre*) have been identified within the genomes of viruses representing each of the major picornavirus genera (*Enterovirus*, *Rhinovirus*, *Aphthovirus*, and *Cardiovirus*) except *Hepatovirus*. Previous efforts to identify a stem-loop structure with *cre* function in hepatitis A virus (HAV), the type species of this genus, by phylogenetic analyses or thermodynamic predictions have not succeeded. However, a region of markedly suppressed synonymous codon variability was identified in alignments of HAV sequences near the 5' end of the 3D<sup>pol</sup>-coding sequence of HAV, consistent with noncoding constraints imposed by an underlying RNA secondary structure. Subsequent MFOLD predictions identified a 110-nucleotide (nt) complex stem-loop in this region with a typical AAACA/G *cre* motif in its top loop. A potentially homologous RNA structure was identified in this region of the avian encephalitis virus genome, despite little nucleotide sequence relatedness between it and HAV. Mutations that disrupted secondary RNA structure or the AAACA/G motif, without altering the amino acid sequence of 3D<sup>pol</sup>, ablated replication of a subgenomic HAV replicon in transfected human hepatoma cells. Replication competence could be rescued by reinsertion of the native 110-nt stem-loop structure (but not an abbreviated 45-nt stem-loop) upstream of the HAV coding sequence in the replicon. These results suggest that this stem-loop is functionally similar to *cre* elements of other picornaviruses and likely involved in templating VPg uridylation as in other picornaviruses, despite its significantly larger size and lower free folding energy.**

Hepatitis A virus (HAV), a member of the family *Picornaviridae*, is a common etiological agent of acute hepatitis in humans. Several characteristics of HAV distinguish it from other picornaviruses, justifying its classification within a separate genus in this family, the genus *Hepatovirus* (10). First, the primary sequence of the positive-sense, single-stranded RNA genome of HAV shares little homology with that of other picornaviruses. Avian encephalitis virus (AEV) is the only exception to this statement, as it bears considerable sequence relatedness to HAV and has been tentatively classified as the second member of this genus. Details of the organization and function of the HAV polyprotein also differ in several respects from the typical LP1P2P3 organization found in other picornaviruses (10). The amino-terminal capsid protein, VP4, is remarkably small, lacks an N-terminal consensus myristoylation signal, and appears to have an alternative function in viral assembly compared to the VP4 proteins of other picornaviruses (24). The largest capsid protein, VP1, also contains a unique C-terminal extension that appears to function in pentamer assembly (3). Congruent with this evidence for a unique assembly pathway for the viral particle, the thermal stability of

the HAV virion far exceeds that of other picornaviruses (7, 20). HAV also lacks a 2A proteinase, and unlike other picornaviruses, the primary proteolytic cleavage of the viral polyprotein at the 2A/2B junction is mediated by the only proteinase expressed by the virus, 3C<sup>pro</sup> (18). HAV also demonstrates several unique biological attributes. Strongly hepatotropic in vivo and restricted in its host range to humans, chimpanzees, and certain New World primates, wild-type (wt) HAV replicates very poorly in cultured cells, demonstrating a protracted replication cycle and typically no evidence of a cytopathic effect in infected cells (10). Recent evidence suggests that the 3A protein of HAV is specifically directed to the outer mitochondrial membrane (28), rather than membranes of the endoplasmic reticulum, as is the case with other picornaviruses, possibly contributing to the low efficiency of replication that typifies this virus.

Despite these striking distinguishing characteristics, HAV shares many features in common with other picornaviruses, particularly in terms of its overall genome organization and apparent replication strategy (10). Like the genomes of other picornaviruses, the HAV genome contains a lengthy 5' untranslated region (5' UTR), followed by the single long open reading frame encoding the polyprotein, a short 3' UTR, and a 3'-polyadenylated tail. The 5' UTR lacks a 5'-terminal m<sup>7</sup>G cap structure and in genomic RNA is covalently linked to a small virus-encoded peptide (3B or VPg) (27). The internal ribosome entry site (IRES) located within the 5' UTR directs the cap-independent translation of the polyprotein, which is cotranslationally processed by the 3C<sup>pro</sup> proteinase to produce structural proteins that comprise the viral capsid and nonstruc-

\* Corresponding author. Mailing address: Center for Hepatitis Research, 4.104 Blocker Medical Research Bldg., University of Texas Medical Branch, 301 University Blvd., Galveston, TX 77555-1073. Phone: (409) 747-7048. Fax: (409) 747-7030. E-mail: smlemon@utmb.edu.

† Present address: University of Texas M. D. Anderson Cancer Center, Houston, TX.

‡ These authors contributed equally to this work.

<sup>∇</sup> Published ahead of print on 6 August 2008.

tural proteins involved in replication of the viral genome (2). Like other picornaviruses, genome replication is a two-stage process, with the input RNA genome first transcribed to produce antisense RNA, which then functions as template for the synthesis of positive-sense progeny genomes (10). 3D<sup>pol</sup>, the RNA-dependent RNA polymerase and catalytic core of the viral replicase, directs the synthesis of both positive- and negative-strand RNA. Although far less well studied than poliovirus (PV), the synthesis of HAV RNA is thought to be primed by VPg-pUpU, the product of 3D<sup>pol</sup>-mediated uridylylation of VPg.

The genomes of all picornaviruses contain RNA replication signals within both the 5'- and 3'-terminal domains. However, the genomes of many other picornaviruses also have been found to contain internally located stem-loop structures that are essential for viral RNA synthesis. First recognized within the capsid coding region of human rhinovirus 14 (HRV-14), its function in viral RNA replication *in vivo* could not be complemented *in trans*, leading to its designation as a *cis*-acting replication element (*cre*) (12). The ability of the *cre* to support viral RNA synthesis is dependent upon both specific RNA structure and certain nucleotides within the loop region (14, 16, 25, 29). On the other hand, it is independent of its position within the genome and of whether its sequence is translated into protein (12). Similar RNA elements have subsequently been identified within the P1 sequence of cardioviruses, the 2C-coding region of PV, the 2A-coding sequence of HRV-2, and the 5' UTR of an aphthovirus (4, 5, 9, 11).

Studies by Paul and colleagues (14, 16, 17) have shown that the PV *cre* functions as the template for VPg (3B) uridylylation through a "slide-back" mechanism catalyzed by 3D<sup>pol</sup> in association with 3CD. The uridylylation of VPg, possibly in the context of 3AB, leads to the production of VPg-pUpU, which serves as the protein primer for new RNA synthesis (8, 15). Extensive mutagenesis of the HRV-14 and PV *cre* revealed a critical conserved AAACA/G motif in the 5' half of the loop sequence that is essential for *cre* function (29). Similar conserved AAACA motifs are present within the loops of the *cre* elements of other picornaviruses and are important for RNA replication (17, 31). Evidence suggests that a *cre* is likely to be present in all picornaviruses but at different positions within the genome in different picornaviruses and with substantial variation in primary nucleotide sequences. To date, however, searches for such an element in the HAV genome have not been productive.

Internal base pairing that creates stem-loops and other RNA structures places constraints on sequence variability in bases required for structure formation. In the hepatitis C virus (HCV) genome, this constraint is manifested by a marked suppression of synonymous codon variability within several evolutionarily conserved stem-loops in the core and NS5B-coding regions that have demonstrated roles in viral replication (13, 26, 32). Discrete RNA structures such as the *cre* in the coding region of human enteroviruses (HEVs) and other viruses that lack other large-scale RNA secondary structures (22) should also create characteristic suppression of synonymous site variability (SSSV), similar to that observed in HCV. Here, we describe the use of independent phylogenetic and thermodynamic methods to scan HAV sequences for covariant sites and associated RNA secondary structures, leading to the

prediction of a conserved stem-loop structure within RNA encoding the 3D<sup>pol</sup> RNA-dependent RNA polymerase. We confirmed the functional importance of this structure by mutagenesis and reverse molecular genetics. We show that this RNA element shares several common features with other picornavirus *cre* elements but is unique in both size and location.

## MATERIALS AND METHODS

**Nucleotide sequences.** The following sequences of human and simian HAV were used for analysis of variability at synonymous sites and for RNA structure determination: NC\_001489, DQ646426, AB258387, AB020567, HPAACG, AB020566, AB020564, AB020565, AB020568, AB020569, AF314208, HPACG, AB300205, AF512536, AF357222, AF485328, HAVCOMPL, HPA18F, AY644670, AY644676, HAVRNAHFS, EF207320, EF406357, EF406359, SHVAGM27, AJ299464, AY974170, and HEA299464. Sequences showing <1% sequence divergence from other published sequences were excluded from analysis. The following nonidentical sequences for AEV were also analyzed for RNA secondary structure: NC\_003990, AJ225173, AY517471, and AY275539. Alignments of HAV and AEV sequences were carried out by identification and alignment of conserved amino acid sequence motifs in the coding region and automated alignment of intervening regions using ClustalW with default settings. Alignments of HEV species A and B viruses corresponded to those used in previous analyses (23).

**Analysis of SSSV.** Synonymous sequence variability was determined by measurement of mean pairwise distances at each codon position in the open reading frames of hepatoviruses and enteroviruses. Variability at each codon was calculated using the program Sequence Scan in the Simmonic sequence editor (21). Mean pairwise synonymous variability was restricted to aligned codons where the translated amino acid was the same. Each pairwise value was normalized by dividing by the degeneracy of the codon, with normalization factors for twofold degenerate sites of 0.5, for threefold degenerate sites of 0.6666, for fourfold degenerate sites of 0.75, and for sixfold degenerate sites of 0.8333. This takes into account the different sequence distances achievable at maximally diverged sites. Variability at each codon position was averaged over a sliding window of 35 codons.

**RNA secondary structure prediction.** Base pairing in the region of the genome showing SSSV was predicted using MFOLD using default settings through the web interface at <http://www.bioinfo.rpi.edu/applications/mfold>. PFOLD analysis used the web interface at <http://www.daimi.au.dk/~compbio/rnafold/>. All programs were run with default settings. Thermodynamic structure predictions were carried out using the program ZIPFOLD on the MFOLD server. Minimum folding energies (MFEs) of native HAV and HEV-B sequences were compared with the mean values generated for 50 control sequences in which sequence order had been scrambled using an algorithm (NDR) that preserves the dinucleotide frequencies of the native sequence, implemented in the Simmonic sequence editor package. MFE differences (MFEDs) were calculated as MFED (%) = [(MFE<sub>native</sub>/MFE<sub>scrambled</sub>) - 1] × 100, where MFE<sub>native</sub> and MFE<sub>scrambled</sub> are the MFEs of native sequences and mean MFEs of the 50 sequence order-randomized controls, respectively.

**Plasmids.** The contribution of RNA structures to replication of the HAV genome was assessed by creating mutations within pHAVLuc, which contains the cDNA of a replication-competent, subgenomic RNA replicon, HAVLuc, in which an in-frame fusion of the firefly luciferase coding sequence replaces all but the 5' 150 and 3' 39 nucleotides (nt) of the P1 region of the HM175/18f genome (30). pHAVLuc-Δ3D is a related replication-incompetent mutant of pHAVLuc, which contains a single base substitution creating a premature termination codon within the 3D<sup>pol</sup> sequence; it is referred to here simply as Δ3D. Mutations disrupting the native RNA sequence of the 3D<sup>pol</sup>-coding region were introduced by QuikChange site-directed mutagenesis (Stratagene). The following oligonucleotides were used for construction of mutations: for MutA, GTTCAATGAA TGTCGTGTCGAAGACCCCTTTTAGAAAGAGTC (+) and GACTCTTCT TAAAAAGGGTCTTCGACACGACATTCATTGAAC (-); for MutB, GCTT TTTAGAAAAAGTCCAATCTACCATCATCATGATAAAAC (+), and GTT TTATCAATGTGATGGTAGATTGGACTTTTCTAAAAAGC (-); and for MutC, GTTCAATGAATGTGGTCTCCAAGACGCTTTTAGAAAGAGTC (+) and GACTCTTCTAAAAAGCGTCTTGGAGACCACATTCATT GAAC (-).

To construct HAV replicons with reinserions of the putative wt and mutated *cre* sequences immediately downstream of the luciferase coding sequence, the SacI site (nt 3006 in the HM175/18f sequence [7, 33]), which was used to fuse the luciferase sequence to sequence encoding the C terminus of VP1 in pHAVLuc,

was eliminated by mutating T<sup>3006</sup> to G (silent base change) to create pHAVLuc\_v.2. A new *SacI* site was then placed at the 3' end of the full-length *cre* sequence by introducing C<sup>6071</sup>G and A<sup>6076</sup>C mutations to create pHAVLuc\_v.3. The sequence between nt 3010 and 5955 (HM175/18f sequence) was deleted by QuikChange mutagenesis to fuse the full-length *cre* sequence in frame to the 3' end of the luciferase sequence in pHAVLuc\_v.3, resulting in pΔHAVLuc\_v.3. Next, pHAVLuc was digested with *SacI* (nt 3006 in the HM175/18f sequence) and *XhoI* (nt 7013), and the small fragment was ligated into the *SacI/XhoI* sites in ΔHAVLuc\_v.3 to create pwt/wt. A similar strategy was used to construct pMutA/MutA, containing two mutated *cre* elements derived from MutA, and pwt/MutA and pMutA/wt. For construction of ps-cre/wt and ps-cre/MutA, the 45-nt *s-cre* RNA segment was introduced into pHAVLuc and pMutA by QuikChange site-directed mutagenesis using the oligonucleotides GTTCAATG GAGCTCTAAGCTTATAAATGGGACTCTTTCTAAAAAGCGTTTTGGA GACCACATTCATGGTACCAATTGGACTTTCC (+) and GTTCAATG GAGCTCTAAGCTTATAAATGGGACTCTTTCTAAAAAGCGTTTTGGA GACCACATTCATGGTACCAATTGGACTTTCC (-). All plasmid regions subjected to PCR mutagenesis were sequenced to ensure that no adventitious mutations were introduced.

**In vitro RNA transcription.** To produce replicon RNA transcripts, plasmids were linearized at the unique *XmaI* restriction site located at the 3' end of the HAV sequence (30). RNA transcripts were synthesized by T7 polymerase-mediated transcription (T7 MEGascript; Ambion). The integrity and yield of the transcribed RNAs were determined by agarose gel electrophoresis.

**Cell culture.** Huh7 human hepatoma cells were grown in Dulbecco's modified Eagle's medium (Gibco/BRL) with 10% fetal bovine serum.

**HAV RNA replication assay.** Huh7 cells were transfected with replicon RNA transcripts. Briefly,  $5 \times 10^6$  cells were electroporated with 10 µg of RNA using a GenePulser II electroporation apparatus (Bio-Rad) with the pulse controller unit set at 1,400 V and 25 µF and maximum resistance. The cells were subsequently seeded into a 12-well plate and cultured in Dulbecco's modified Eagle's medium with 10% fetal bovine serum at 37°C until processing for luciferase assays. Cell lysates were harvested by the addition of 100 µl of passive lysis buffer (Promega) to each well and stored at -20°C until assayed for enzymatic activity. Luciferase activity was quantified using the luciferase assay system (Promega) as described by the supplier, with results determined using a TD-20/20 luminometer (Turner Designs).

## RESULTS

**Prediction of an RNA structure within the N-terminal region of the HAV 3D<sup>pol</sup>-coding sequence.** To investigate whether the RNA secondary structure of a previously defined *cre* in the 2C region of HEVs (5) could be identified by suppression of variability at synonymous sites, alignments of complete coding sequences of species A and B isolates were scanned for variability at each codon (Fig. 1A). Synonymous site variability was extremely high throughout the coding sequences of both viruses, except for a short region in 2C. The area of maximum SSSV coincided precisely with the previously reported stem-loop forming the *cre* (Fig. 1A). The analysis was repeated using 28 complete coding sequences of human and simian HAV variants (after identical or near-identical sequences, including sequences of multiple HM175 strain variants, had been removed) (Fig. 1B). Although both the nucleotide and inferred amino acid sequence diversity within HAV was substantially lower than that observed within HEV species, a single discrete region of SSSV was observed at the start of the 3D<sup>pol</sup>-coding sequence (Fig. 1B).

To investigate whether the 5' end of the 3D<sup>pol</sup>-coding sequence contained RNA secondary structure that would account for the observed SSSV, sequences from each of the HAV sequences between positions 5501 and 6500 (HAV positions are numbered according to the wt HM175 virus sequence, NC\_001489, unless otherwise noted) were analyzed by MFOLD to derive minimum free energy predicted structures

(Fig. 2). A conserved stem-loop was predicted for the sequences between positions 5948 and 6057, whereas base pairings either side of the stem-loop were variable in different strains of HAV (data not shown). Despite several nucleotide substitutions, sequences of the more divergent simian HAV strains also formed an RNA structure that was similar in pairing and shape to the human HAV structure (Fig. 2). PFOLD predicted the same consensus secondary structure using the alignment of human and simian HAV sequences (data not shown).

The existence of thermodynamically favored RNA secondary structure formation in the *cre* region of HEVs and a region of suppressed synonymous variability in HAV sequences was investigated by comparison of MFEs of native sequences with those of sequence order-randomized controls (Table 1). HEV-B sequences showed high MFED values (+23% for the 100-base fragment that included the *cre*), consistent with a major contribution of sequence order to the secondary structure in this region. However, MFEs for native and scrambled HAV sequences were approximately the same (MFEDs of -1 to -2%). Possible reasons for the absence of detectable stability differences between native sequence of HAV and controls are discussed below.

Although substantially larger, the predicted stem-loop structure in the HAV 3D<sup>pol</sup> region showed several features that are common to *cre* elements present in other picornavirus genera. These include a large open top loop (18 bases, larger than the 14 bases in HRV-14, HRV-2, and HEV [4, 5, 12] and 15 bases in foot-and-mouth disease virus [11]), as well as the presence of the AAACA/G motif in the 5' half of the loop that has been shown to play a templating role in uridylation of VPg (16, 29). Importantly, the region of predicted internal base pairing showed a higher frequency of invariant or minimally variable codons than did flanking regions (Fig. 3). For the 35 codons in the stem-loop for which there are potential synonymous alternatives, 20 were invariant (57%) and nine showed variability of <0.33 (26%). These frequencies are significantly higher than variability in flanking regions (14 invariant, 25 minimally variable from a total of 107 codons; 13% and 23%, respectively,  $P < 0.0000013$  by the  $\chi^2$  test). Importantly, the codons comprising the AAACA/G motif and adjacent sequence within the 5' half of the loop segment were absolutely invariant in their sequence (Fig. 3).

To further investigate the phylogenetic conservation of the 3D<sup>pol</sup> stem-loop in HAV, we examined sequences from the distantly related AEV. AEV is tentatively classified within the genus *Hepatovirus*, but recent work has revealed that its IRES structure differs significantly from that of HAV and is actually closer to that of HCV (1). This and other considerations have prompted a reconsideration of its taxonomic classification. The lack of sequence diversity in published complete genome sequences of AEV precluded a meaningful analysis of SSSV for this virus. However, while coding sequences from AEV are highly divergent from those of HAV, a defensible alignment of HAV and AEV sequences was possible in the region of the genome around the start of the 3D<sup>pol</sup>-coding sequence (Fig. 4). MFOLD analysis of the region homologous to positions 5501 to 6500 of HAV demonstrated a stem-loop structure in a position equivalent to that in HAV, in the virtual absence of nucleotide sequence similarity between viruses (Fig. 2). Although the actual pairings in the stem



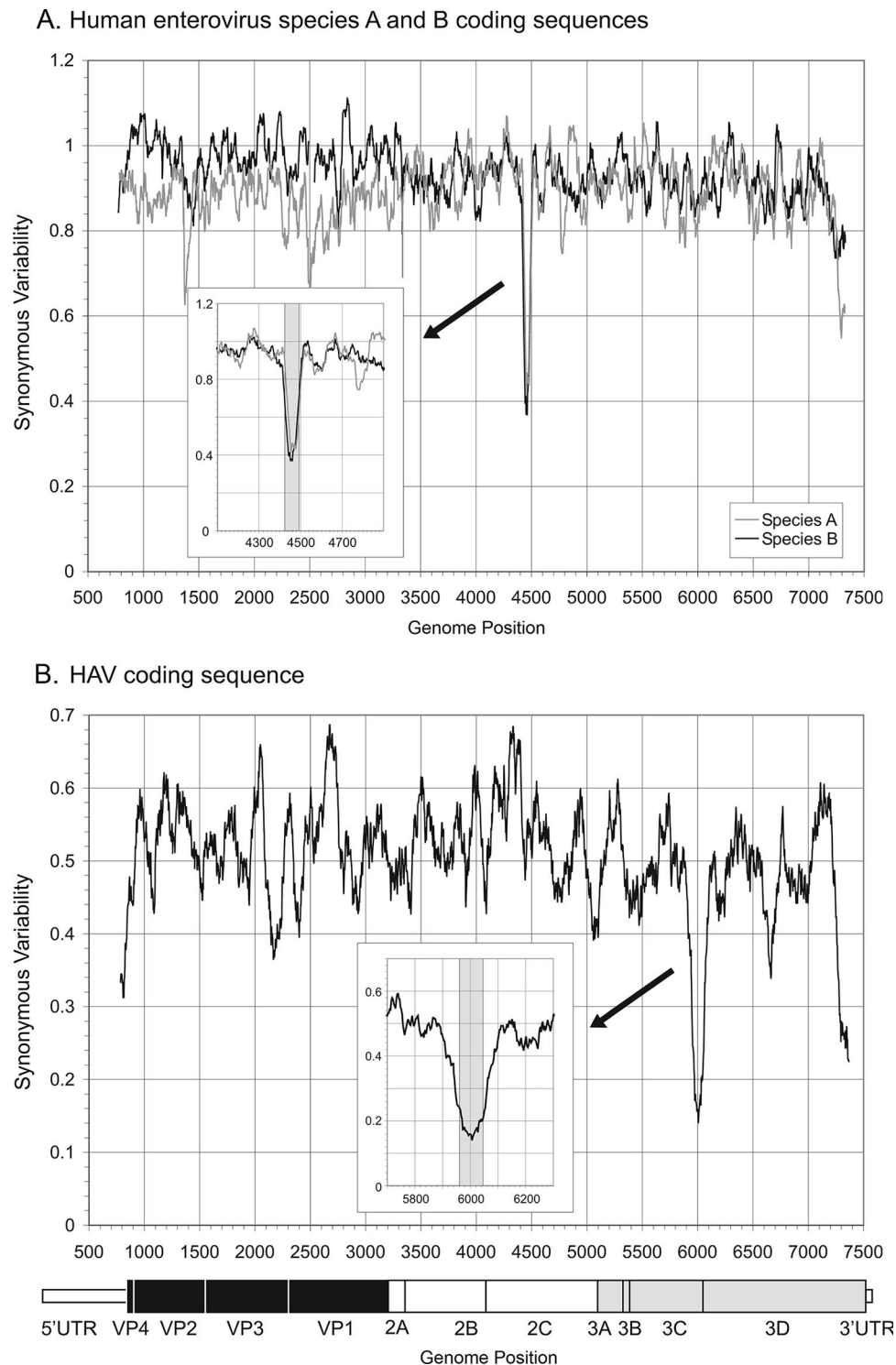


FIG. 1. Scans searching for SSSV in the coding region of the nucleotide sequences of HEVs and human hepatoviruses. (A) SSSV scan of HEV species A and B sequences. The inset graph shows a defined region of SSSV on a larger *x* scale, with the region of known RNA base pairing associated with the enterovirus *cre* superimposed (gray shading, positions 4436 to 4496 in the POL3L27 sequence). (B) Similar SSSV scan of the human HAV sequence. Nucleotide positions are numbered according to the wt HM175 virus sequence (NC\_001489).

differed considerably between HAV and AEV, the AEV structure formed a very similar unpaired terminal loop of 16 bases, with the AAACG sequence placed in a position homologous to that in the HAV loop.

**RNA structure within the 3D<sup>pol</sup>-coding region is essential for efficient HAV RNA replication.** To assess whether the putative stem-loop structure identified within the hepatovirus 3D<sup>pol</sup>-coding region has functional importance, possibly equiv-

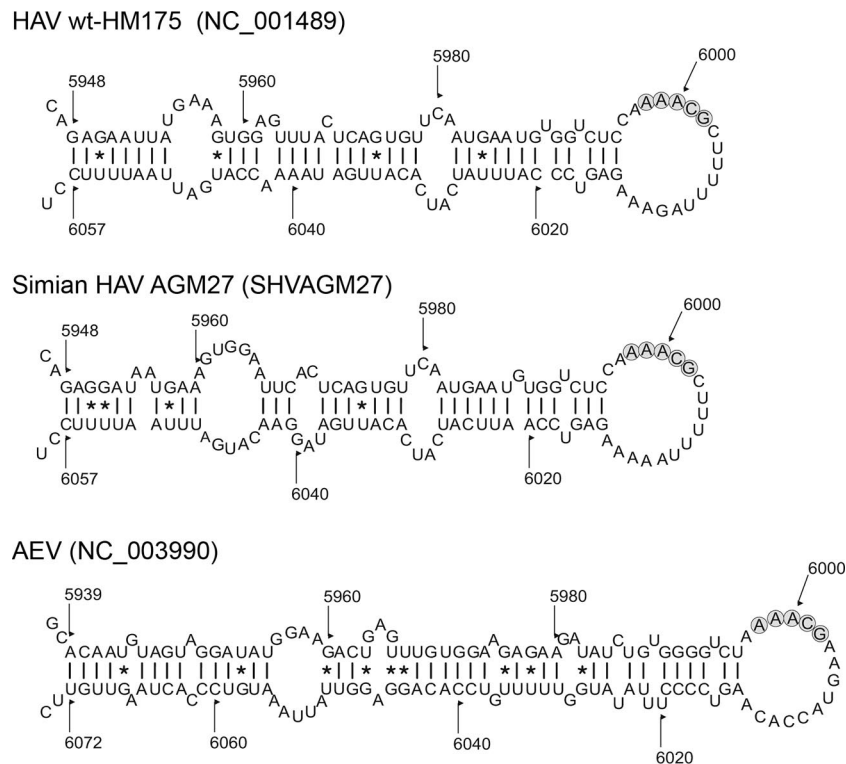


FIG. 2. Consensus MFOLD predicted RNA structures for human and simian HAVs and AEV in the 3D<sup>pol</sup>-coding region showing significant SSSV. Numbering for each structure corresponds to that for human HAV (NC\_001489).

alent to that of the *cre* of other picornaviruses, we constructed two mutants (MutA and MutB) within the background of a replication-competent, subgenomic RNA replicon, HAVLuc, in which an in-frame fusion of the firefly luciferase coding sequence replaces most of the P1, capsid coding region of the genome of a rapidly replicating, cell culture-adapted variant of HAV (Fig. 5A) (30). To create MutA, we introduced five silent, third-base mutations in the 5' half of the stem-loop, including two that disrupted predicted base pairs near the top of the stem and two within the AAACG sequence in the loop (Fig. 5B). MutB contained four silent mutations within the 3' sequence of the upper part of the stem (Fig. 5B). Thus, the multiple nucleotide substitutions created in these mutants disrupt the AAACG/A motif and/or predicted base pairing within the helix, without altering the amino acid sequence of the 3D<sup>pol</sup> RNA-dependent RNA polymerase. The MFOLD algorithm predicts that the mutations in either of these mutants

significantly disrupt the RNA structure predicted within this segment of the wt genome (data not shown).

To evaluate the impact of these mutations on HAV RNA replication, we transcribed RNA in vitro from both wt and mutant constructs and transfected these RNAs into human hepatoma (Huh7) cells that support HAV replication. As a negative control, we also transfected Huh7 cells with a related, replication-defective RNA, HAVLuc-Δ3D (hereafter referred to as Δ3D), in which translation of 3D<sup>pol</sup> is prematurely terminated (30). Transfection of this RNA leads to an initial burst of luciferase expression, detectable at 24 h posttransfection, but no subsequent increase in luciferase activity at 48 or 72 h. This early luciferase expression represents translation of the input RNA, which is degraded following transfection. In contrast, the replication-competent wt RNA (HAVLuc), while generating a similar level of luciferase activity at 24 h, shows a sustained, almost 10-fold increase at later time points (Fig. 5C). Interestingly, although none of the nucleotide substitutions that we engineered into MutA or MutB altered the amino acid sequence of the polypeptide, both RNAs behaved like the 3D<sup>pol</sup> mutant, generating luciferase only at 24 h and with little or no luciferase activity detectable at 48 or 72 h posttransfection. The similar levels of luciferase activity generated by the mutants and the wt HAVLuc at 24 h posttransfection suggest that the nucleotide substitutions engineered into MutA and MutB do not reduce the efficiency with which the HAV IRES directs the translation of the polypeptide in these RNAs. However, the lack of a subsequent increase in luciferase expression in cells transfected with either mutant indicates that these nucleotide substitutions are severely detrimental to HAV

TABLE 1. MFEs of sequences flanking the HAV versus HEV-B *cre*<sup>a</sup>

<i>cre</i> region	Position (bp)			Length (bp)	MFE		MFED (%)
	Beginning	End	Midpoint		Native	Control	
HAV	5952	6051	6002	100	-15.4	-15.7	-1.9
	5855	6151	6003	300	-56.7	-57.2	-0.9
HEV-B	4401	4500	4451	100	-22.8	-18.6	22.6
	4305	4604	4455	300	-71.8	-67.2	6.8

<sup>a</sup> Comparison of MFEs of 100- and 200-nt segments centered on the *cre* in the human HAV and HEV-B genomes with those of sequence order-scrambled controls. The MFED quantifies the contribution of sequence order to RNA secondary structure.



FIG. 3. Distribution of synonymous site variability among codons forming the 3D<sup>pol</sup> stem-loop and flanking sequences of HAV. The 38 codons contributing to the complex stem-loop structure shown in Fig. 2 (nt 5946 to 6059) are separated from flanking codons in the center of the figure.

RNA replication. Since the amino acid sequences of the polyproteins of MutA and MutB were unchanged from those of the wt genome, these results suggest that the predicted RNA structure in the 3D<sup>pol</sup>-coding sequence is critically important for HAV RNA replication and that it may in fact function as a *cre*.

To further evaluate this possibility, we created a third mutant, MutC, containing only a single base alteration, A<sup>5999</sup> to G (Fig. 5B). This mutation does not alter the MFOLD-predicted secondary structure in this region of the genome, nor the amino acid sequence of 3D<sup>pol</sup>, but changes the AAACG sequence to AGACG, thus knocking out the AAACA/G motif. Since the second adenosine in this motif is critically important to the slide-back mechanism underlying the 3D<sup>pol</sup>-mediated uridylation of VPg (16, 17), this mutation would be expected to be lethal to replication, if in fact the predicted stem-loop functions as a *cre*. Consistent with this hypothesis, the MutC

RNA also failed to replicate following transfection into Huh7 cells (Fig. 5C).

**Functional rescue of MutA by the full-length but not a short, 45-nt RNA element.** An interesting characteristic of the picornavirus *cre* is its ability to function in supporting viral RNA replication in a manner that is independent of its position within the genome (5, 12). We thus attempted to rescue the replication competence of MutA by inserting the entire 110-nt stem-loop sequence (Fig. 6A, f-*cre*, nt 5946 to 6059) downstream of the HAV coding sequence, at unique restriction sites engineered into MutA between the stop codon terminating translation of 3D<sup>pol</sup> and the 3' UTR. This replicon RNA showed no evidence of replication following transfection into Huh7 cells (data not shown). However, since an RNA pseudoknot has been proposed previously to form around the junction of the 3D<sup>pol</sup>-coding sequence and the 3' UTR (6), it is likely that the insertion of the f-*cre* sequence at this position

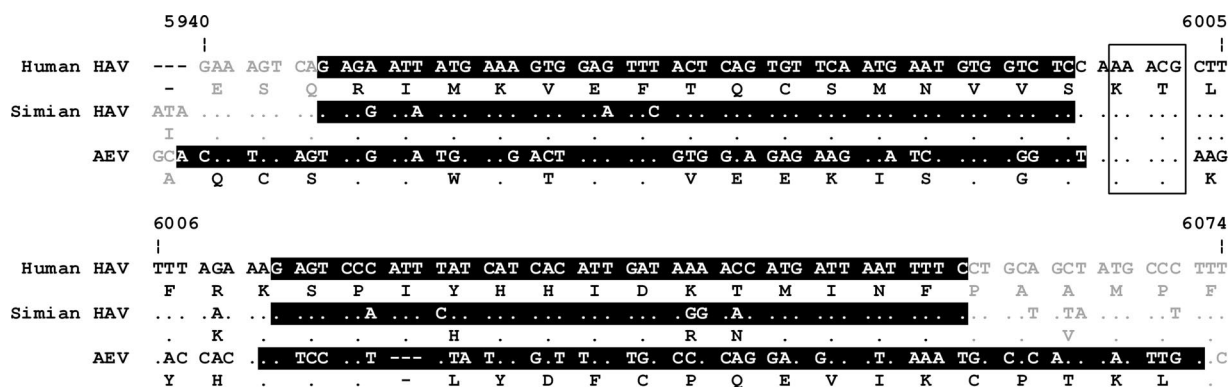


FIG. 4. Alignment of nucleotide sequences based on inferred amino acid sequences of human and simian HAVs and of AEV. Nucleotides within the 3D<sup>pol</sup> stem-loops of HAV and AEV are shown in bold, with bases within the stem regions of the structure shown in inverse. The AAACA/G motif in the loop region (AAACG in HAV) is boxed. Sequences used are those shown in Fig. 2.





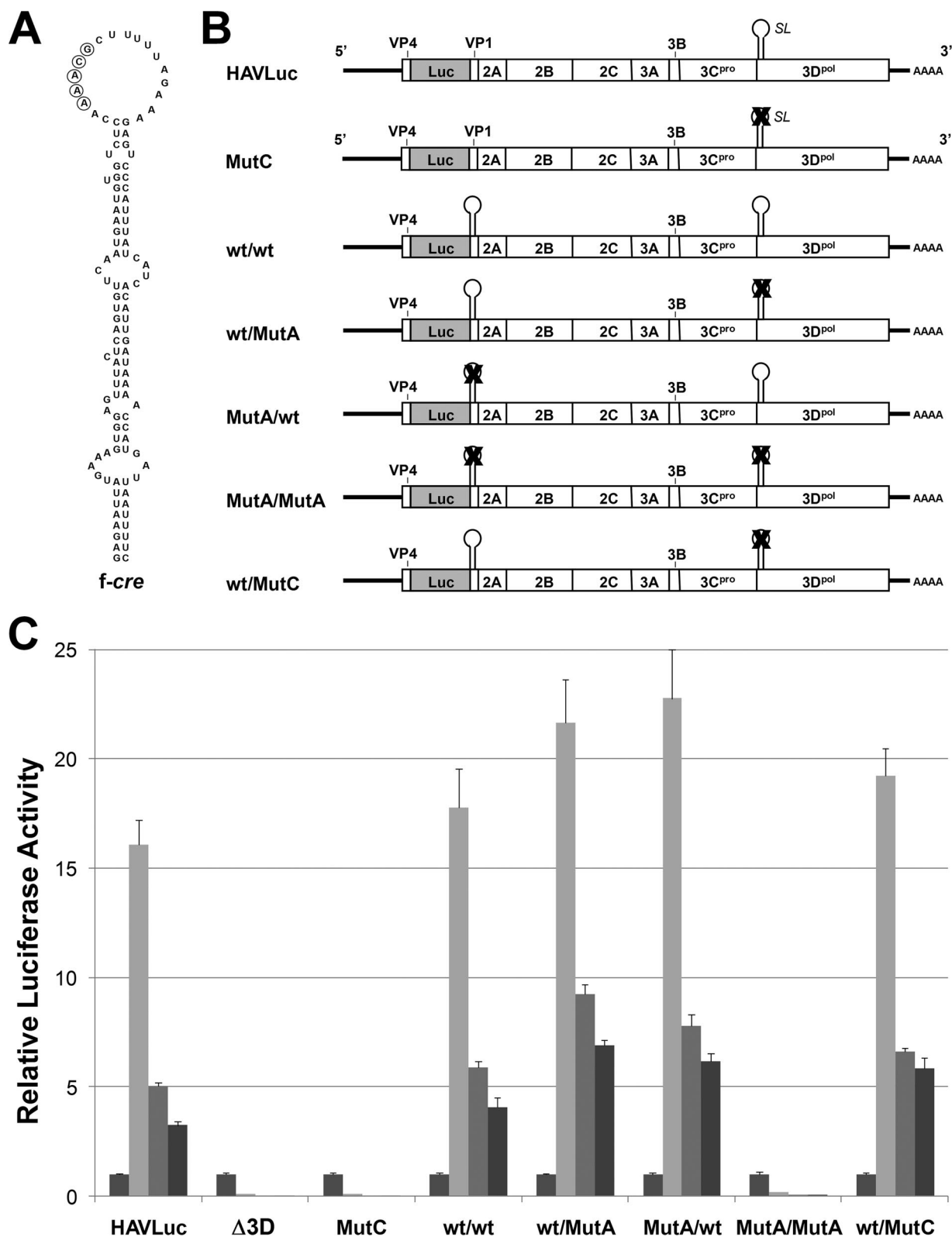


FIG. 6. Replication competence is rescued by reinsertion of the 3D<sup>pol</sup> stem-loop fused downstream of the luciferase sequence in the MutA replicon. (A) The full-length, wt 110-nt *f-cre* sequence inserted downstream of the luciferase sequence in the constructs shown in panel B. (B) Schematic showing the insertion of the wt *f-cre* stem-loop downstream of the luciferase sequence in the wt/wt, wt/MutA, and wt/MutC replicons. MutA/wt and MutA/MutA contain a similarly inserted, mutated *f-cre* sequence (mutations identical to MutA as in Fig. 5B) in the background of HAVLuc and MutA, respectively. (C) Relative luciferase activities expressed by HAVLuc, Δ3D, MutC (Fig. 5), wt/wt, wt/MutA, MutA/wt, MutA/MutA, and wt/MutC at 24, 48, 84, and 106 h (left to right, respectively, within each set of bars) post-transfection of Huh7 cells. The results shown are the averages of nine independent cultures of transfected cells; error bars indicate the standard deviations.

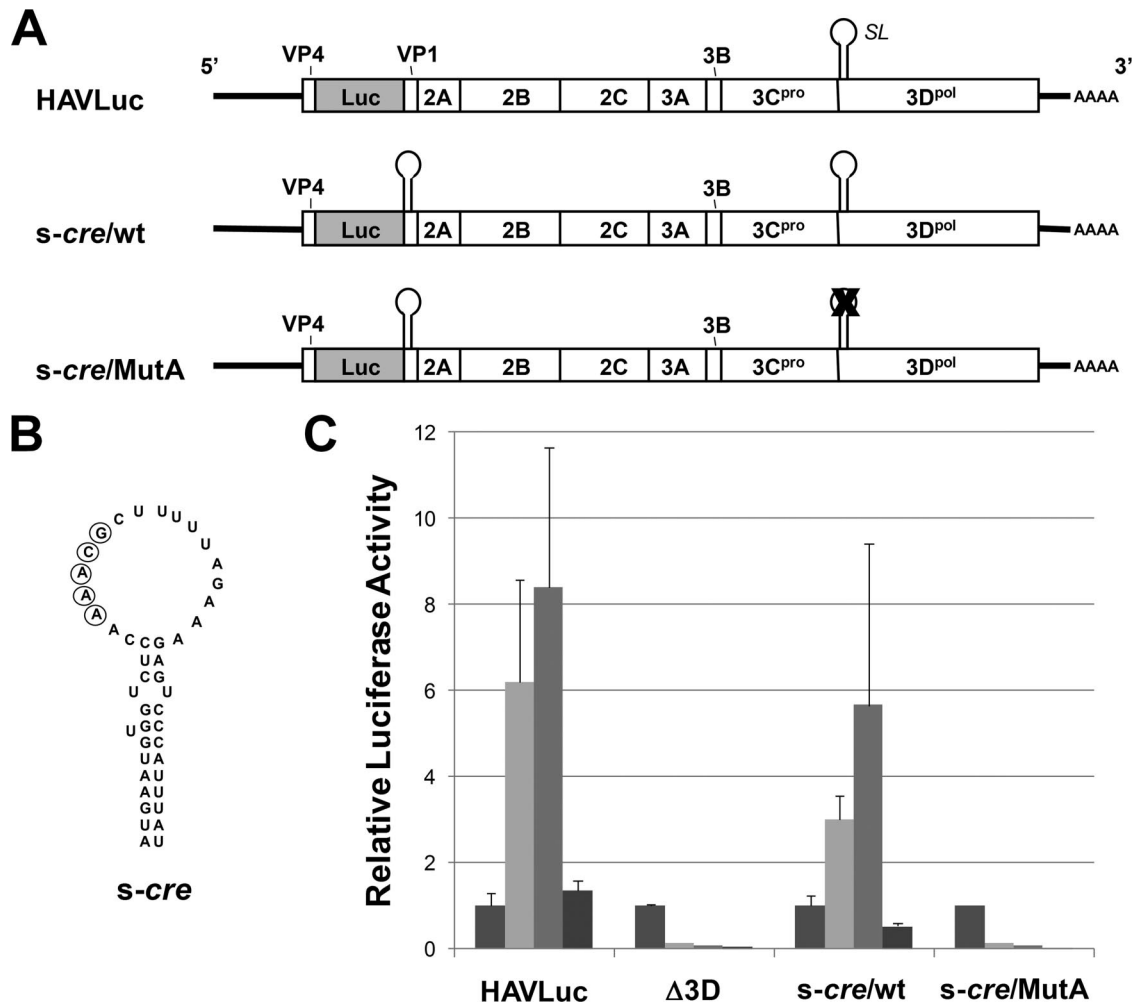


FIG. 7. An abbreviated 45-nt *cre* sequence (*s-crc*) is incapable of rescuing the replication competence of MutA. (A) Schematic showing in-frame insertion of the 45-nt *s-crc* sequence between the luciferase and residual VP1 coding sequence in the background HAVLuc (*s-crc/wt*) or MutA (*s-crc/MutA*). (B) The 45-nt *s-crc* sequence inserted within the constructs shown in panel A. (C) Relative luciferase activities expressed by HAVLuc, Δ3D, *s-crc/wt*, and *s-crc/MutA* at 24, 48, 72, and 96 h post-transfection of Huh7 cells (bars are shaded as in Fig. 5C). See the legend to Fig. 5C for details.

the insertion was well tolerated in the background of HAVLuc (Fig. 7C, *s-crc/wt*). Although further studies were not undertaken, these data indicate that HAV RNA replication requires a substantially lengthier *cre* than does that of other picornaviruses, possibly as long as the entire *f-crc* sequence (Fig. 6A). This may be related to the relatively low free folding energy of the HAV *cre* compared to that of other picornavirus *cre* stem-loop structures, as discussed below.

## DISCUSSION

Conserved internal stem-loop structures that function as templates for the uridylation of VPg and that are thus essential for RNA replication have been identified previously in representative viruses from most of the major pathogenic picornavirus genera. Such a *cre* element has not, however, been identified previously within the genome of hepatoviruses. Previous studies have shown that the sequence spanning the VP4 to 2A coding regions of HAV can be deleted without loss of

RNA replication competence, indicating that there is not an essential *cre* element residing within the P1 sequence of HAV. Thus, it was of interest to find a large, stable RNA structure existing within the P3 segment encoding the viral RNA-dependent RNA polymerase, 3D<sup>pol</sup>. This structure conforms to the general requirements of a picornavirus *cre* element, including an AAACA/G motif appropriately placed within the top loop and a location-independent function essential for viral RNA replication that requires preservation of both RNA secondary structure and sequence within the loop (5, 12, 17, 29). The AAACA/G motif is particularly important for *cre* function, as tandem adenosine residues within it template the uridylation of VPg by a slide-back mechanism (16, 17, 29). Thus, the fact of the ablation of the critical second adenosine residue within the motif in MutC is particularly informative. This single base change altered neither MFOLD-predicted secondary structure nor amino acid sequence within 3D<sup>pol</sup> but nonetheless completely eliminated replication of the HAVLuc replicon (Fig. 5C). Further biochemical studies will be required to confirm a

role in VPg uridylylation for this novel HAV stem-loop structure.

Thus, taken together, our experimental data strongly suggest that the conserved RNA structure that we have identified near the 5' end of the 3D<sup>pol</sup>-coding sequence is in fact the HAV *cre*. Importantly, it appears to be present in all hepatoviruses, including both human and simian HAVs, as well as the distantly related avian virus AEV (Fig. 2). However, in contrast to other recognized picornavirus *cre* elements, such as the enterovirus and rhinovirus *cre* elements, it is a significantly larger structure. The minimal functional HRV-14 *cre* requires no more than a top loop of 14 nt extending from a helical segment of only 9 bp in order to support RNA replication in vivo or VPg uridylylation in cell-free reactions (29). Thus, only 33 nt of sequence is required to support RNA replication. In contrast, the structure that we identified in the HAV 3D<sup>pol</sup>-coding region is 110 nt in length and contains a top loop of 18 nt with a much lengthier stem segment comprising 35 bp, interrupted by four internal loops and two 1-nt bulges (Fig. 2). The high-resolution nuclear magnetic resonance structure of the HRV-14 *cre* indicates that the 14-nt loop segment adopts a specific fold that derives stability from base stacking interactions (25). It is not possible to predict whether a similar structure might be adopted by the larger loop segment of the HAV *cre*. While further studies will be needed to determine the minimal functional HAV *cre*, a 45-nt segment containing the top loop and immediately adjacent 12-bp helical segment was inadequate to support HAV RNA replication (Fig. 7). To some extent, this may reflect the low free energy on folding associated with this RNA structure and the low G+C content of hepatovirus/AEV genome sequences (38% and 42%, respectively).

We identified this RNA structure by combined phylogenetic and thermodynamic predictive strategies. We first identified a region within the protein coding sequence of the HAV genome that displayed marked suppression of synonymous variability among codons (Fig. 1B). Since a similar SSSV scan of HEV species revealed a unique low-variability signal that aligned precisely with the previously identified enterovirus *cre* (Fig. 1A), the low synonymous site variability signature identified within the hepatovirus genome may be similarly indicative of an equivalent *cre* element. This low synonymous variation within base-paired regions of the *cre* elements arises from the need to conserve nucleotide sequence in order to maintain both regional RNA secondary structure and the amino acid sequence of the protein encoded by the RNA. MFOLD analyses of the region identified by the SSSV scan revealed a large stem-loop structure that is conserved across members of the genus *Hepatovirus* and led to the mutational analysis described above. Interestingly, the HAV *cre* was "invisible" using standard MFOLD folding free energy scanning. There was no detectable difference between the folding free energy of the HAV coding segment containing the *cre* and that of sequence content order-scrambled controls (Table 1), even though it was not difficult to identify the enterovirus *cre* with this method. Both MFEs (per base) and the arithmetical difference between the MFEs of native and scrambled control sequences (MFEDs) are much lower for 100- and 300-nt segments centered on the HAV than for the HEV species B *cre*.

The reasons underlying these differences in the HAV *cre* and the *cre* elements of other picornaviruses are not clear. One

consideration is that there may be specific differences between functional RNA structures in viral genomes with different G+C contents and that the HAV *cre* may have to be larger (compared to the G+C-rich enterovirus and rhinovirus *cre* elements) because of a larger proportion of A:U pairings within the duplex region of the A+U-rich HAV structure. Twenty-five of the 35 bp in the *cre* sequence of HAV are either A:U or G:U (Fig. 2). However, only four of the nine base pairs in the minimal HRV-14 *cre* are G:C pairings (29), compared to 5 of 12 in the top part of the stem in the HAV *s-cre* sequence that was not capable of supporting RNA replication (Fig. 7). This region of the HRV-14 and PV *cre* appears to be important for recognition by 3CD and 3D<sup>pol</sup>, which is important in establishing the complex that leads to uridylylation of VPg (19). The fact that a *cre* appears to exist within the HAV genome, as shown here, suggests that the general scheme of VPg uridylylation resulting in the protein primer for RNA synthesis that has been established for PV most likely holds true for HAV as well. However, there are likely to be important differences in the details of this process, given the remarkably different size and folding free energy of the HAV *cre* stem-loop structure as revealed in the studies described here, the apparent targeting of the HAV 3A (and likely 3AB) protein to mitochondrial rather than endoplasmic reticulum membranes, and differences in the biophysical properties of the VPg protein of HAV from those of other picornaviruses (27, 28).

Whatever the basis may be for the longer *cre* loop and overall greater size of the HAV *cre*, the identification of this essential replication element enhances our understanding of the molecular virology of HAV, an important but relatively ignored human pathogen. Like almost all other aspects of this fascinating virus, the HAV *cre* shows both significant similarities and substantial important differences in comparison to related aspects of other well-studied picornavirus genera.

#### ACKNOWLEDGMENTS

We are grateful for expert assistance provided by Yinghong Ma and Jeremy Yates.

This work was supported in part by a grant from the National Institutes of Health, U19-AI40035. Yan Yang was supported by a McLaughlin Postdoctoral Fellowship.

#### REFERENCES

1. Bakhshesh, M., E. Gropelli, M. M. Willcocks, E. Royall, G. J. Belsham, and L. O. Roberts. 2008. The picornavirus avian encephalomyelitis virus possesses a hepatitis C virus-like internal ribosome entry site element. *J. Virol.* **82**:1993–2003.
2. Brown, E. A., S. P. Day, R. W. Jansen, and S. M. Lemon. 1991. The 5' nontranslated region of hepatitis A virus: secondary structure and elements required for translation in vitro. *J. Virol.* **65**:5828–5838.
3. Cohen, L., D. Benichou, and A. Martin. 2002. Analysis of deletion mutants indicates that the 2A polypeptide of hepatitis A virus participates in virion morphogenesis. *J. Virol.* **76**:7495–7505.
4. Gerber, K., E. Wimmer, and A. V. Paul. 2001. Biochemical and genetic studies of the initiation of human rhinovirus 2 RNA replication: identification of a *cis*-replicating element in the coding sequence of 2A<sup>pro</sup>. *J. Virol.* **75**:10979–10990.
5. Goodfellow, L., Y. Chaudhry, A. Richardson, J. Meredith, J. W. Almond, W. Barclay, and D. J. Evans. 2000. Identification of a *cis*-acting replication element within the poliovirus coding region. *J. Virol.* **74**:4590–4600.
6. Kusov, Y., M. Weitz, G. Dollenmeier, V. Gauss-Muller, and G. Siegl. 1996. RNA-protein interactions at the 3' end of the hepatitis A virus RNA. *J. Virol.* **70**:1890–1897.
7. Lemon, S. M., P. C. Murphy, P. A. Shields, L.-H. Ping, S. M. Feinstone, T. Cromeans, and R. W. Jansen. 1991. Antigenic and genetic variation in cytopathic hepatitis A virus variants arising during persistent infection: evidence for genetic recombination. *J. Virol.* **65**:2056–2065.

8. Liu, Y., D. Franco, A. V. Paul, and E. Wimmer. 2007. Tyrosine 3 of poliovirus terminal peptide VPg(3B) has an essential function in RNA replication in the context of its precursor protein, 3AB. *J. Virol.* **81**:5669–5684.
9. Lobert, P. E., N. Escriou, J. Ruelle, and T. Michiels. 1999. A coding RNA sequence acts as a replication signal in cardioviruses. *Proc. Natl. Acad. Sci. USA* **96**:11560–11565.
10. Martin, A., and S. M. Lemon. 2002. The molecular biology of hepatitis A virus, p. 23–50. *In* J.-H. Ou (ed.), *Hepatitis viruses*. Kluwer Academic Publishers, Norwell, MA.
11. Mason, P. W., S. V. Bezborodova, and T. M. Henry. 2002. Identification and characterization of a *cis*-acting replication element (*cre*) adjacent to the internal ribosome entry site of foot-and-mouth disease virus. *J. Virol.* **76**:9686–9694.
12. McKnight, K. L., and S. M. Lemon. 1998. The rhinovirus type 14 genome contains an internally located RNA structure that is required for viral replication. *RNA* **4**:1569–1584.
13. McMullan, L. K., A. Grakoui, M. J. Evans, K. Mihalik, M. Puig, A. D. Branch, S. M. Feinstone, and C. M. Rice. 2007. Evidence for a functional RNA element in the hepatitis C virus core gene. *Proc. Natl. Acad. Sci. USA* **104**:2879–2884.
14. Paul, A. V., E. Rieder, D. W. Kim, J. H. Van Boom, and E. Wimmer. 2000. Identification of an RNA hairpin in poliovirus RNA that serves as the primary template in the *in vitro* uridylation of VPg. *J. Virol.* **74**:10359–10370.
15. Paul, A. V., J. H. Van Boom, D. Filippov, and E. Wimmer. 1998. Protein-primed RNA synthesis by purified poliovirus RNA polymerase. *Nature* **393**:280–284.
16. Paul, A. V., J. Yin, J. Mugavero, E. Rieder, Y. Liu, and E. Wimmer. 2003. A “slide-back” mechanism for the initiation of protein-primed RNA synthesis by the RNA polymerase of poliovirus. *J. Biol. Chem.* **278**:43951–43960.
17. Rieder, E., A. V. Paul, D. W. Kim, J. H. van Boom, and E. Wimmer. 2000. Genetic and biochemical studies of poliovirus *cis*-acting replication element *cre* in relation to VPg uridylation. *J. Virol.* **74**:10371–10380.
18. Schultheiss, T., Y. Y. Kusov, and V. Gauss-Müller. 1994. Proteinase 3C of hepatitis A virus (HAV) cleaves the HAV polyprotein P2–P3 at all sites including VP1/2A and 2A/2B. *Virology* **198**:275–281.
19. Shen, M., Q. Wang, Y. Yang, H. B. Pathak, J. J. Arnold, C. Castro, S. M. Lemon, and C. E. Cameron. 2007. Human rhinovirus type 14 gain-of-function mutants for oril utilization define residues of 3C(D) and 3Dpol that contribute to assembly and stability of the picornavirus VPg uridylation complex. *J. Virol.* **81**:12485–12495.
20. Siegl, G., M. Weitz, and G. Kronauer. 1984. Stability of hepatitis A virus. *Intervirology* **22**:218–226.
21. Simmonds, P., I. Karakasiliotis, D. Bailey, Y. Chaudhry, D. J. Evans, and I. G. Goodfellow. 2008. Bioinformatic and functional analysis of RNA secondary structure elements among different genera of human and animal caliciviruses. *Nucleic Acids Res.* doi:10.1093/nar/gkn096.
22. Simmonds, P., A. Tuplin, and D. J. Evans. 2004. Detection of genome-scale ordered RNA structure (GORS) in genomes of positive-stranded RNA viruses: implications for virus evolution and host persistence. *RNA* **10**:1337–1351.
23. Simmonds, P., and J. Welch. 2006. Frequency and dynamics of recombination within different species of human enteroviruses. *J. Virol.* **80**:483–493.
24. Tesar, M., X.-Y. Jia, D. F. Summers, and E. Ehrenfeld. 1993. Analysis of a potential myristoylation site in hepatitis A virus capsid protein VP4. *Virology* **194**:616–626.
25. Thiviyanathan, V., Y. Yang, K. Kaluarachchi, R. Rijnbrand, D. G. Gorenstein, and S. M. Lemon. 2004. High-resolution structure of a picornaviral internal *cis*-acting RNA replication element (*cre*). *Proc. Natl. Acad. Sci. USA* **101**:12688–12693.
26. Tuplin, A., D. J. Evans, and P. Simmonds. 2004. Detailed mapping of RNA secondary structures in core and NS5B-encoding region sequences of hepatitis C virus by RNase cleavage and novel bioinformatic prediction methods. *J. Gen. Virol.* **85**:3037–3047.
27. Weitz, M., B. M. Baroudy, W. L. Maloy, J. R. Ticehurst, and R. H. Purcell. 1986. Detection of a genome-linked protein (VPg) of hepatitis A virus and its comparison with other picornaviral VPgs. *J. Virol.* **60**:124–130.
28. Yang, Y., Y. Liang, L. Qu, Z. Chen, M. Yi, K. Li, and S. M. Lemon. 2007. Disruption of innate immunity due to mitochondrial targeting of a picornaviral protease precursor. *Proc. Natl. Acad. Sci. USA* **104**:7253–7258.
29. Yang, Y., R. Rijnbrand, K. L. McKnight, E. Wimmer, A. Paul, A. Martin, and S. M. Lemon. 2002. Sequence requirements for viral RNA replication and VPg uridylation directed by the internal *cis*-acting replication element (*cre*) of human rhinovirus type 14. *J. Virol.* **76**:7485–7494.
30. Yi, M., and S. M. Lemon. 2002. Replication of subgenomic hepatitis A virus RNAs expressing firefly luciferase is enhanced by mutations associated with adaptation of virus to growth in cultured cells. *J. Virol.* **76**:1171–1180.
31. Yin, J., A. V. Paul, E. Wimmer, and E. Rieder. 2003. Functional dissection of a poliovirus *cis*-acting replication element [PV-*cre*(2C)]: analysis of single- and dual-*cre* viral genomes and proteins that bind specifically to PV-*cre* RNA. *J. Virol.* **77**:5152–5166.
32. You, S., D. D. Stump, A. D. Branch, and C. M. Rice. 2004. A *cis*-acting replication element in the sequence encoding the NS5B RNA-dependent RNA polymerase is required for hepatitis C virus RNA replication. *J. Virol.* **78**:1352–1366.
33. Zhang, H. C., S. F. Chao, L. H. Ping, K. Grace, B. Clarke, and S. M. Lemon. 1995. An infectious cDNA clone of a cytopathic hepatitis A virus: genomic regions associated with rapid replication and cytopathic effect. *Virology* **212**:686–697.